North-Holland

NON-REPETITIVE WORDS RELATIVE TO A REWRITING SYSTEM

Arturo CARPI

Dipartimento di Matematica ed Applicazioni dell'Università di Napoli, Italy, and L.I.T.P., Paris, France

Aldo de LUCA

Dipartimento di Matematica dell'Università di Roma "La Sapienza" and Istituto di Cibernetica del C.N.R., Arco Felice, Napoli, Italy

1. Introduction

The construction of very long non-repetitive words on finite alphabets constitutes a subject of interest in combinatorics on words (cf. [1, 8]). A fundamental result of Thue [10] states that a free monoid over a three letter alphabet has infinitely many words containing no identical adjacent blocks of letters (*square-free words*). On the contrary, as a consequence of the iteration lemma, one has that in any infinite context-free language there are words containing a square.

In this paper we introduce the notion of a *square-free word relative to a rewriting system*. It is a word which generates, by means of the given rewriting system, a set of square-free words. We deal with two decision problems: to decide whether a given word is square-free relative to a given rewriting system (*square-free word problem*) and to decide whether there are infinitely many square-free words relative to a rewriting system (*infinite square-free word problem*).

Symmetric rewriting systems (i.e. Thue systems) are also presentations of monoids, so that one can consider the problem of deciding whether the presented monoid contains an infinite number of square-free elements (equivalently, whether there exists an infinite quotient monoid in which $x^2=0$ for all $x \neq 1$). We show that if one restricts oneself to considering symmetric rewriting systems without productions of the type $(1, v)$ with v a non-empty word, then this problem is equivalent to the infinite square-free word problem.

Afterwards we consider some particular kinds of rewriting systems with just one rule or one symmetric couple of rules and context-free systems. We show that the infinite square-free word problem for these three classes of rewriting systems is decidable as well as the square-free word problem for one rule systems and for context-free systems.

In the last two sections systems of *semi-commutations* are taken into account. Our motivation is to seek a generalization of the results [4], in which the infinite square-free word problem for partially commutative free monoids is shown to be

0304-3975/90/$3.50

decidable. In Section 7 we consider also the notion of an *overlap-free word relative to a rewriting system* π and we show that in the case of semi-commutations one can decide whether there exist infinitely many overlap-free words relative to π.

We conclude by quoting some problems which arise from our investigation: Does there exist a rewriting system whose square-free word problem is undecidable? Is the infinite square-free word problem (for arbitrary rewriting systems) decidable? Finally, is the infinite square-free word problem for systems of semi-commutations decidable?

2. Preliminaries

Let A be a finite alphabet. We denote by A^* and A^+ the free monoid and the free semigroup generated by A, respectively. The elements of A are called *letters* and those of A^* *words*. The *empty word* is denoted by 1. For any word $w \in A^*$, $|w|$ will denote its *length* and alph(w) the set of all letters occurring at least once in w.

An *infinite word* on A is a sequence of letters $\boldsymbol{a} = a_0a_1 \ldots a_n \ldots$ $(a_n \in A, n \geq 0)$. All the words $a_ma_{m+1} \ldots a_n$ $(n \geq m \geq 0)$ are said to be *factors* of $\boldsymbol{a}$.

A *rewriting* (*or semi-Thue*) *system* on A is a finite subset π of the cartesian product $A^* \times A^*$. The pairs $(p, q) \in \pi$ are called *productions*. We denote by $\Rightarrow$ the regular closure of π and by $\Rightarrow^*$ the reflexive and transitive closure of $\Rightarrow$. In other terms, one has $f \Rightarrow g$ $(f, g \in A^*)$ if and only if

$$f = hpk, \qquad g = hqk$$

for some $h, k, p, q \in A^*$ such that $(p, q) \in \pi$, while one has $f \Rightarrow^* g$ if and only if either $f = g$ or there exist $f_1, f_2, \ldots, f_n \in A^*$ $(n \geq 2)$ such that $f = f_1$, $g = f_n$ and $f_i \Rightarrow f_{i+1}$, for all $i = 1, 2, \ldots, n-1$.

For any $u \in A^*$ the set

$$L_u(\pi) = \{w \in A^* \mid u \overset{*}{\Rightarrow} w\}$$

is said to be the *language generated by u* (by means of π).

A rewriting system π on A is said to be *context-free* if $\pi \subseteq A \times A^*$. A *system* of *semi-commutations* is a rewriting system of the form

$$\pi = \{(ab, ba) \mid (a, b) \in \theta\},$$

where θ is a reflexive relation on A.

We introduce now the notion of square-free word relative to a rewriting system. We recall that a word $w \in A^*$ is said to be *square-free* if one cannot factorize w as $w = rsst$ with $r, s, t \in A^*$ and s non-empty. The set of all square-free words on the alphabet A will be denoted by $L_2(A)$. We say that w is *square-free relative to the rewriting system* π if

$$L_w(\pi) \subseteq L_2(A),$$

that is, if the language generated by w contains only square-free words. The set of square-free words relative to π will be denoted by $L_2(A, \pi)$. We remark that $L_2(A, \pi)$

is stable for factors, as one can easily verify in view of the regularity of the relation $\Rightarrow^*$. An infinite word $\boldsymbol{a}$ on A is said to be *square-free relative to* π if all its factors are square-free relative to π. One can consider the following two problems.

Square-free word problem. Given a rewriting system π on an alphabet A, decide for any $v \in A^*$ whether v is square-free relative to π.

Infinite square-free word problem. Decide, for any rewriting system π in a given class, whether there exists an infinite square-free word relative to π.

We remark that, by König's Lemma, the existence of an infinite square-free word relative to π is equivalent to the infiniteness of $L_2(A, \pi)$.

If the rewriting system π contains a production $(1, v)$ with $v \in A^+$, then $L_2(A, \pi)$ is empty since $w \Rightarrow^* wvv$, for all $w \in A^*$. For this reason, in the sequel we consider only rewriting systems $\pi \subseteq A^+ \times A^*$.

We conclude this section by introducing the notion of an inalterable word. Let π be a rewriting system on the alphabet A. We say that a word v is *inalterable* (relative to π) if $L_v(\pi) = \{v\}$. In other terms, v is inalterable if one cannot factorize v as $v = hpk$ with $h, p, k \in A^*$ and $(p, q) \in \pi$ for some $q \in A^*$, $q \neq p$. An infinite word $\boldsymbol{a}$ is *inalterable* (relative to π) if all its factors are inalterable. It is evident that an inalterable square-free word is necessarily square-free relative to π, and therefore if there exist arbitrary long inalterable square-free words, then $L_2(A, \pi)$ is infinite.

In general, the converse is not true. For instance, if one sets $A = \{a, b, c, d\}$, $\pi = \{(a, b), (b, a)\}$, then any square-free word of length longer than 3 is not inalterable; nevertheless $L_2(A, \pi)$ is infinite.

3. Thue systems

Let us suppose that the rewriting system π is *symmetric*, i.e. $\pi^{-1} \subseteq \pi$. In this case π is sometimes called a *Thue system*. The relation $\Rightarrow^*$ is a congruence and therefore one can consider the quotient monoid

$$M(A, \pi) = A^*/\overset{*}{\Rightarrow}.$$

An element $m \in M(A, \pi)$ is said to be *square-free* if one cannot factorize m as $m = rsst$ with $r, s, t \in M(A, \pi)$ and $s \neq i$, where i is the identity of $M(A, \pi)$. It is evident that the square-free elements of $M(A, \pi)$ are the congruence classes (mod $\Rightarrow^*$) of the words belonging to $L_2(A, \pi)$. So, if $L_2(A, \pi)$ is finite, then $M(A, \pi)$ contains only a finite number of square-free elements. The converse is less trivial: indeed, one could think that $L_2(A, \pi)$ can contain a finite number of congruence classes (mod $\Rightarrow^*$), some of which are infinite. We shall show that this cannot happen.

Proposition 3.1. *Let π be a Thue system on the alphabet A. $L_2(A, \pi)$ is infinite if and only if $M(A, \pi)$ contains infinitely many square-free elements.*

Proof. We have to show that if $M(A, \pi)$ contains only a finite number of square-free elements, then $L_2(A, \pi)$ is finite. Let us consider the congruence $\equiv$ on A^* generated by

$$\pi \cup \{(u, v) \in A^* \times A^* \mid u, v \notin L_2(A, \pi)\}.$$

It is evident that $A^* \backslash L_2(A, \pi)$ is a congruence class (mod $\equiv$) while $\equiv$ coincides with $\Rightarrow^*$ on the set $L_2(A, \pi)$. So, if we suppose that $M(A, \pi)$ contains a finite number r of square-free elements, then $\equiv$ has finite index $r+1$. We deduce by the Myhill theorem of automata theory that $L_2(A, \pi)$ is a rational language. Now $L_2(A, \pi)$ must be finite: otherwise, in fact, by the iteration lemma, it would contain some non-square-free words. □

We remark explicitly that Proposition 3.1 is not true if π contains a production $(1, v)$ with $v \neq 1$. For instance if $A = \{a, b, c, d\}$ and $\pi = \{(1, a), (a, 1)\}$ then $L_2(A, \pi)$ is empty, while all the square-free words on the alphabet $\{b, c, d\}$ are projected into square-free elements of $M(A, \pi)$.

4. One-rule systems

A *one-rule Thue system* is a rewriting system on an alphabet A of the form $\pi = \{(u, v), (v, u)\}$ $(u, v \in A^*)$. It is unknown whether there exist one-rule Thue systems with an undecidable word-problem (cf. [3] for partial results on this subject). On the contrary, we shall show that the square-free word problem for one-rule Thue systems is decidable.

We begin with the following.

Proposition 4.1. *Let A be an alphabet, L a subset of A^+ and π the rewriting system on A defined by*

$$\pi = \{(u, v) \mid u, v \in L, u \neq v\}. \tag{4.1}$$

Factorize any $w \in A^*$ in the form $w = r_0x_1r_1 \ldots x_nr_n$ with $n \geq 0$, $x_i \in L$ $(1 \leq i \leq n)$, $r_j \in A^* \backslash A^*LA^*$ $(0 \leq j \leq n)$ and set

$$T = \{r_0y_1r_1 \ldots y_nr_n \mid y_i \in L, 1 \leq i \leq n\}.$$

Then one has $w \in L_2(A, \pi)$ if and only if

(i) *T is stable for $\Rightarrow$,*

(ii) *$T \subseteq L_2(A)$.*

Proof. We suppose $w \in L_2(A, \pi)$ and show that T is stable for $\Rightarrow$.

If $w_1 \in T$ and $w_1 \Rightarrow w_2$, then one has

$$w_1 = r_0y_1r_1 \ldots y_nr_n = rz_1s, \qquad w_2 = rz_2s$$

with y_i, $z_j \in L$ $(1 \leq i \leq n, j = 1, 2)$.

At least one of the following conditions is verified:

$$(1)\qquad z_1=z'z'',\qquad z',z''\neq 1,\qquad rz'=r_0y_1r_1\dots r_{i-1},\qquad z''s=y_ir_i\dots r_n,$$

$$(2)\qquad z_1=z'z'',\qquad z',z''\neq 1,\qquad rz'=r_0y_1r_1\dots y_i,\qquad z''s=r_iy_{i+1}\dots r_n,$$

$$(3)\qquad y_i=pz_1q,\qquad r=r_0y_1r_1\dots r_{i-1}p,\qquad s=qr_iy_{i+1}\dots r_n$$

($1\leq i\leq n$). Situations (1) and (2) correspond to the case that z_1 is broken by a parsing-line of the factorization $w_1=r_0y_1r_1\dots y_nr_n$. Situation (3) corresponds to the case that z_1 appears as a factor of y_i. (We recall that z_1 cannot be a factor of any r_j.)

In case (1) one has

$$w\overset{*}{\Rightarrow} r_0y_1r_1\dots y_{i-1}r_{i-1}z_1r_iy_{i+1}\dots r_n=r(z')^2z''r_iy_{i+1}\dots r_n.$$

In case (2) one has

$$w\overset{*}{\Rightarrow} r_0y_1r_1\dots y_{i-1}r_{i-1}z_1r_iy_{i+1}\dots r_n=r_0y_1\dots r_{i-1}z'(z'')^2s.$$

In both cases one obtains a contradiction, since $w\in L_2(A,\pi)$. Hence, condition (3) is necessarily verified. We deduce

$$w\overset{*}{\Rightarrow} w_1=rz_1s\overset{*}{\Rightarrow} ry_is=rpz_1qs\overset{*}{\Rightarrow} rpy_iqs=rp^2z_1q^2s$$

and, from the assumption that $w\in L_2(A,\pi)$, we derive $p=q=1$. Hence, one has $z_1=y_i$ and

$$w_2=r_0y_1r_1\dots r_{i-1}z_2r_iy_{i+1}\dots r_n\in T.$$

If $w\in L_2(A,\pi)$, then one has

$$T\subseteq L_w(\pi)\subseteq L_2(A)$$

and therefore (ii) is verified.

Conversely, let us suppose that (i), (ii) are verified. Since $w\in T\subseteq L_w(\pi)$ and $L_w(\pi)$ is the minimal subset of A^* containing w and stable for $\Rightarrow$, we deduce $L_w(\pi)=T\subseteq L_2(A)$ and therefore $w\in L_2(A,\pi)$. □

Suppose that L is finite. Then, in view of the finiteness of T, one can effectively decide whether conditions (i) and (ii) are verified. Thus, the previous proposition furnishes an algorithm to solve the square-free word problem for finite Thue systems of the form (4.1). In particular, we have the following.

Corollary 4.2. *The square-free word problem for one-rule Thue systems is recursively solvable.*

A similar result can be obtained for *one-rule semi-Thue systems*, i.e. systems of the form $\pi=\{(u,v)\}$ with $u,v\in A^*$. More precisely, one has the following.

Proposition 4.3. *Let $\pi = \{(u, v)\}$ be a one-rule semi-Thue system on an alphabet A, such that $u \neq 1$. For any $w \in A^*$ factorize w as $w = r_0 u r_1 u r_2 \dots u r_n$ ($n \geq 0$, $r_j \in A^* \backslash A^* u A^*$, $0 \leq j \leq n$) and set*

$$T = \{r_0 y_1 r_1 y_2 r_2 \dots y_n r_n \mid y_i \in \{u, v\}, 1 \leq i \leq n\}.$$

Then one has $w \in L_2(A, \pi)$ if and only if

(i) *T is stable for $\Rightarrow$,*

(ii) *$T \subseteq L_2(A)$.*

In this case the word $w' = r_0 v r_1 v r_2 \dots v r_n$ is inalterable.

The proof of this proposition is similar to that of Proposition 4.1 and is therefore omitted.

Proposition 4.3 not only gives an algorithm to solve the square-free word problem for a one-rule semi-Thue system π but it also assures that $L_2(A, \pi)$ is infinite if and only if there exists an infinite inalterable square-free word. We shall see that one can effectively verify whether such a word exists. The infinite square-free word problem for one rule semi-Thue systems is therefore decidable.

Proposition 4.4. *Let $\pi = \{(u, v)\}$ be a one-rule semi-Thue system on an alphabet A. One has that*

(i) *if* Card$(A) \geq 4$, *then $L_2(A, \pi)$ is infinite if and only if $u \neq 1$.*

(ii) *if* Card$(A) = 3$, *then $L_2(A, \pi)$ is infinite if and only if u contains two occurrences of the same letter.*

(iii) *if* Card$(A) \leq 2$, *then $L_2(A, \pi)$ is finite.*

Proof. (iii) is trivial since any word of length 4 on a 2-letter alphabet contains a square. To prove (i), it is sufficient to remark that if $a \in$ alph(u), then any square-free word on the alphabet $A \backslash \{a\}$ is inalterable.

Suppose now that Card$(A) = 3$ and that u contains two occurrences of the same letter. Then, by renaming the letters of A we can limit ourselves to considering the case that $A = \{a, b, c\}$ and u contains at least one of the factors *aa*, *cbc*, *bcab*. As remarked in [10] there exists an infinite square-free word $\boldsymbol{a}$ on A which does not contain the factors *cbc* and *cabac*. Obviously *aa* is not a factor of $\boldsymbol{a}$. Suppose that *bcab* is a factor of $\boldsymbol{a}$: by deleting, possibly, the first letter of $\boldsymbol{a}$ we can suppose that $\boldsymbol{a}$ does not begin with b and therefore $\boldsymbol{a}$ contains a factor *xbcabyz* ($x, y, z \in A$). Since for any $x, y, z \in A$ the word *xbcabyz* contains a square or an occurrence of *cbc* or *cabac*, we obtain a contradiction. We deduce that $\boldsymbol{a}$ is inalterable and therefore $L_2(A, \pi)$ is infinite.

Finally, we consider the case in which A is a 3-letter alphabet and u does not contain two occurrences of the same letter. In this case, as remarked by Thue, all the infinite square-free words contain u as a factor. In view of Proposition 4.3, we deduce that $L_2(A, \pi)$ is finite. $\square$

We can give a partial extension of Proposition 4.4 to the case of one-rule Thue systems.

Proposition 4.5. *Let* $\pi = \{(u, v), (v, u)\}$ *be a one-rule Thue system on an alphabet* A, *with* $u \neq 1$, $v \neq 1$, $\mathrm{Card}(A) \geqslant 4$. *Then* $L_2(A, \pi)$ *is infinite.*

Proof. We distinguish three cases.

Case 1: $\mathrm{alph}(u) \cap \mathrm{alph}(v) \neq \emptyset$. If $a \in \mathrm{alph}(u) \cap \mathrm{alph}(v)$ then any word on the alphabet $A \backslash \{a\}$ is inalterable. Since $\mathrm{Card}(A \backslash \{a\}) \geqslant 3$, there are arbitrarily long inalterable square-free words and therefore $L_2(A, \pi)$ is infinite.

Case 2: $\mathrm{alph}(u) \cap \mathrm{alph}(v) = \emptyset$, $\min\{|u|, |v|\} \geqslant 2$. By renaming the letters of A, we can reduce ourselves to the case that A contains four distinct letters x, x^{-1}, y, y^{-1} such that xx or xx^{-1} is a factor of u and yy or yy^{-1} is a factor of v. As shown in [7], there are arbitrarily long square-free words on the alphabet $\{x, x^{-1}, y, y^{-1}\}$ which do not contain the factors xx^{-1}, yy^{-1}. Evidently all these words are inalterable and therefore $L_2(A, \pi)$ is infinite.

Case 3: $\mathrm{alph}(u) \cap \mathrm{alph}(v) = \emptyset$, $\min\{|u|, |v|\} = 1$. Without loss of generality, we can suppose $u \in A$, $v \in (A \backslash \{u\})^*$. One can easily verify that for all $w \in (A \backslash \{u\})^*$, $x \in A^*$, one has $w \Rightarrow^* x$ if and only if $\varphi(x) = w$, where $\varphi : A^* \to (A \backslash \{u\})^*$ is the morphism induced by

$$\varphi(u) = v, \qquad \varphi(a) = a, \quad \text{for all } a \in A \backslash \{u\}.$$

We deduce that if w is square-free, then x is square-free whenever $w \Rightarrow^* x$ and therefore $w \in L_2(A, \pi)$. Since there are arbitrarily long square-free words on the alphabet $A \backslash \{u\}$, we conclude that $L_2(A, \pi)$ is infinite. □

We are not able to give an explicit description of those one-rule Thue systems π on a 3-letter alphabet $A = \{a, b, c\}$, such that $L_2(A, \pi)$ is infinite. We can observe that there are cases in which $L_2(A, \pi)$ is finite (as, for instance, when $\pi = \{(ab, ba), (ba, ab)\}$ and cases in which $L_2(A, \pi)$ is infinite (for instance, when $\pi = \{(aba, cbc), (cbc, aba)\}$).

5. Context-free systems

Let π be a context-free rewriting system on the alphabet A. Then, for any $w \in A^*$, $L_w(\pi)$ is a context-free language and therefore one can decide whether $L_w(\pi)$ is infinite. If $L_w(\pi)$ is infinite, then by the iteration lemma one obtains

$$L_w(\pi) \not\subseteq L_2(A).$$

If, on the contrary, $L_w(\pi)$ is finite, then one can effectively verify whether $L_w(\pi) \subseteq L_2(A)$. So one has the following.

Proposition 5.1. *The square-free word problem for a system of context-free productions is recursively solvable.*

Also the infinite square-free word problem for systems of context-free productions is recursively solvable: an algorithm to solve it is given by the following proposition.

Proposition 5.2. *Let $\pi \subseteq A \times A^*$ be a context-free rewriting system. $L_2(A, \pi)$ is infinite if and only if there exist three letters $a, b, c \in A$ such that each two of $L_a(\pi)$, $L_b(\pi)$ and $L_c(\pi)$ are disjoint subsets of A.*

Proof. Suppose that $L_2(A, \pi)$ is infinite and fix $w \in L_2(A, \pi)$ such that $|w| \geqq 4\,\mathrm{Card}(A)+4$. As remarked previously, $L_w(\pi)$ is finite. Let v_1 be a word of maximal length in $L_w(\pi)$. Then one has

$$L_a(\pi) \subseteq A \cup \{1\}, \quad \text{for all } a \in \mathrm{alph}(v_1).$$

By deleting in v_1 all the letters a such that $a \stackrel{*}{\Rightarrow} 1$, we obtain a word v_2 such that

$$v_1 \stackrel{*}{\Rightarrow} v_2, \qquad L_a(\pi) \subseteq A, \quad \text{for all } a \in \mathrm{alph}(v_2).$$

Moreover one has

$$v_2 = a_1 a_2 \dots a_t, \qquad v_1 = x_0 a_1 x_1 a_2 x_2 \dots a_t x_t$$

for some $a_i \in A$, $x_j \in A^*$ $(0 \leqslant j \leqslant t, 1 \leqslant i \leqslant t)$ such that $x_j \stackrel{*}{\Rightarrow} 1$; x_j cannot contain two occurrences of the same letter $b \in A$, otherwise one would have $x_j \stackrel{*}{\Rightarrow} bb$. So, one must have $|x_j| \leqslant \mathrm{Card}(A)$. We deduce

$$4\,\mathrm{Card}(A)+4 \leqslant |w| \leqslant |v_1| < (t+1)(\mathrm{Card}(A)+1)$$

and hence $|v_2| = t > 3$.

Now, let v_3 be a word of $L_{v_2}(\pi)$ such that $\mathrm{alph}(v_3)$ has minimal cardinality. One has

$$L_a(\pi) \subseteq A, \quad \text{for all } a \in \mathrm{alph}(v_3), \qquad |v_3| = |v_2| > 3$$

and, since v_3 is square-free, $\mathrm{Card}(\mathrm{alph}(v_3)) \geqslant 3$. Moreover whenever $a, b \in \mathrm{alph}(v_3)$ and $a \neq b$, one has $L_a(\pi) \cap L_b(\pi) = \emptyset$. Otherwise, indeed, by substituting in v_3 all the occurrences of a and b with a letter $c \in L_a(\pi) \cap L_b(\pi)$, one would obtain a word of $L_{v_2}(\pi)$ on a smaller alphabet.

Thus we have proved that if $L_2(A, \pi)$ is infinite, then the condition required in the statement is verified. Conversely, it is evident that if a, b, c are such that $L_a(\pi)$, $L_b(\pi)$ and $L_c(\pi)$ are disjoint subsets of A then any square-free word on the alphabet $\{a, b, c\}$ is an element of $L_2(A, \pi)$ and therefore $L_2(A, \pi)$ is infinite. $\square$

By the preceding proof, one can derive that if $L_2(A, \pi)$ is finite, then the maximal length of its words is linearly bounded by $\mathrm{Card}(A)$ independently of π. We have seen, in fact, that if there exists $w \in L_2(A, \pi)$ such that $|w| \geqslant 4\,\mathrm{Card}(A)+4$, then one can find a, b, $c \in A$ verifying the condition required in the statement and therefore $L_2(A, \pi)$ is infinite. By refining this argument, one could show that if $L_2(A, \pi)$ is finite then the maximal length of its words is not longer than $4\,\mathrm{Card}(A)-5$. This bound is optimal: indeed if one sets $A=\{a_1, a_2, \ldots, a_n\}$, $\pi=\{(a_i, 1) \mid 1 \leq i \leq n-2\}$ and

$$w = a_1a_2 \ldots a_{n-2}a_{n-1}a_1a_2 \ldots a_{n-2}a_na_1a_2 \ldots a_{n-2}a_{n-1}a_1a_2 \ldots a_{n-2},$$

then $L_2(A, \pi)$ is finite by Proposition 5.2 and w is a word of length $4n-5$ which is square-free relative to π.

6. Semi-commutations

Let θ be a reflexive relation on the alphabet A. We can then consider the system of semi-commutations

$$\pi = \{(ab, ba) \mid (a, b) \in \theta\}.$$

π is length-preserving (i.e. one has $|u|=|v|$ whenever $u \Rightarrow v$, $u, v \in A^*$) and therefore the language generated by any word $w \in A^*$ is finite. We deduce that the square-free word problem for systems of semi-commutations is recursively solvable.

If θ is symmetric, then π is a Thue system and therefore one can consider the monoid $M(A, \pi)$, which is said to be the *partially commutative free monoid* on A (relative to θ). These objects have been studied in different contexts by several authors (cf. [5, 2, 6] and references therein). An effective characterization of the partially commutative free monoids containing an infinite number of square-free elements has been given by the authors in [4]. The graph of the complementary relation $\bar{\theta}=A\times A\backslash\theta$ of θ is said to be the *non-commutation graph.* With this terminology we have the following.

Proposition 6.1. *A partially commutative free monoid $M(A, \pi)$ contains infinitely many square-free elements if and only if its non-commutation graph contains at least one of the subgraphs shown in Fig.* 1.

Corollary 6.2. *The infinite square-free word problem for symmetric systems of semi-commutations is recursively solvable.*

The situation is more complicated when non-symmetric systems of semi-commutations are taken into account. We will show that one can always limit oneself to the case where the non-commutation graph is strongly connected.

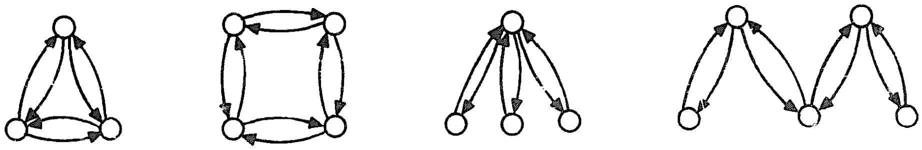

Fig. 1.

Let θ be a reflexive relation on A. We say that A is *reducible* (relative to θ) if one can decompose A as

$$A = A_1 \cup A_2 \quad (A_1, A_2 \neq \emptyset, A_1 \cap A_2 = \emptyset)$$

with $A_1 \times A_2 \subseteq \theta$. We denote by θ_i $(i = 1, 2)$ the restrictions of θ to A_i and set

$$\pi_i = \{(ab, ba) \mid (a, b) \in \theta_i\}.$$

Thus, if $\pi = \{(ab, ba) \mid (a, b) \in \theta\}$ one has $\pi = \pi_1 \cup \pi_2 \cup \{(ab, ba) \mid a \in A_1, b \in A_2\}$. We now have the following.

Proposition 6.3. *Suppose that A is reducible (relative to θ), i.e. $A = A_1 \cup A_2$ $(A_1, A_2 \neq \emptyset, A_1 \cap A_2 = \emptyset)$ and $A_1 \times A_2 \subseteq \theta$. If $L_2(A, \pi)$ is infinite then at least one of the sets $L_2(A_1, \pi_1)$ and $L_2(A_2, \pi_2)$ is infinite.*

Proof. Let us first show that if $w \in A^*$ then $w \Rightarrow^* vu$ for some $v \in A_2^*$, $u \in A_1^*$. The proof is by induction on the length of w.

The result is trivial if $w = 1$. Let us suppose $w \neq 1$ and write

$$w = fx, \quad f \in A^*, x \in A.$$

By the induction hypothesis, $f \Rightarrow^* f_2 f_1$ for some $f_i \in A_i^*$ $(i = 1, 2)$ and therefore

$$w = fx \overset{*}{\Rightarrow} f_2 f_1 x.$$

If $x \in A_1$, the result is achieved. Otherwise we have $f_1 x \Rightarrow^* x f_1$ and then $f_2 f_1 x \Rightarrow^* f_2 x f_1$ with $f_2 x \in A_2^*$.

Suppose now that $w \in L_2(A, \pi)$. Then one has $u \in L_2(A_1, \pi_1)$ and $v \in L_2(A_2, \pi_2)$. If $L_2(A, \pi)$ is infinite, then the length of w can be chosen arbitrarily large. Since $|w| = |u| + |v|$, we deduce that at least one of the sets $L_2(A_1, \pi_1)$ and $L_2(A_2, \pi_2)$ is infinite. $\square$

As a consequence of the previous proposition we obtain the following.

Corollary 6.4. *Let π be a system of semi-commutations on the alphabet A such that $L_2(A, \pi)$ is infinite. Then there exists $B \subseteq A$ such that the restriction to B of the non-commutation graph is strongly connected and $L_2(B, \pi|_B)$ is infinite.*

Proof. The proof is by induction on Card(A). If the non-commutation graph of π is strongly connected, then there is nothing to prove. If, on the contrary, it is not connected, then the alphabet A is reducible. So, by the preceding proposition, there exists $C \subset A$ such that $L_2(C, \pi|_C)$ is infinite and Card$(C) <$ Card(A). Then the conclusion follows from the induction hypothesis. $\square$

In Section 3, we have seen that if π is symmetric, then the infiniteness of $L_2(A, \pi)$ is equivalent to the existence of an infinite number of square-free elements in the monoid $M(A, \pi)$. Let us now suppose that π is a non-symmetric system of semi-commutations. We can then consider the maximal symmetric relation π_2 included in π. We shall see that the infiniteness of $L_2(A, \pi)$ in this case is equivalent to the existence of infinitely many square-free elements in a particular subset of the monoid $M(A, \pi_2)$.

Proposition 6.5. *Let θ be a relation on A. Set*

$$\pi = \{(ab, ba) \mid (a, b) \in \theta\},$$

$$\pi_1 = \pi \setminus \pi^{-1} = \{(ab, ba) \mid (a, b) \in \theta, (b, a) \notin \theta\},$$

$$\pi_2 = \pi \cap \pi^{-1} = \{(ab, ba) \mid (a, b) \in \theta, (b, a) \in \theta\}$$

and denote by J the ideal of $M(A, \pi_2)$ generated by the projection of the set $S = \{ab \mid (ab, ba) \in \pi_1\}$. Then $L_2(A, \pi)$ is infinite if and only if there are infinitely many square-free elements in the set $M(A, \pi_2) \setminus J$.

In other words, $L_2(A, \pi)$ is infinite if and only if there exist arbitrarily long words $w \in A^*$ such that $L_w(\pi_2)$ avoids the squares and the set S. Another formulation of the previous property is the following: consider the congruence $\equiv$ on $A^* \cup \{0\}$ generated by

$$\pi_2 \cup \{(ab, 0) \mid ab \in S\} \cup \{(x^2, 0) \mid x \in A^+\}.$$

Then $L_2(A, \pi)$ is infinite if and only if the monoid $(A^* \cup \{0\})/\equiv$ is infinite; indeed the words of A^* such that $L_w(\pi_2)$ avoids the squares and S are exactly the words which are projected into a non-null element of $(A^* \cup \{0\})/\equiv$.

The proof of Proposition 6.5 uses the following.

Lemma 6.6. *In the hypotheses of Proposition 6.5, for any $w \in A^*$ there exists $v \in L_w(\pi)$ such that*

$$L_v(\pi) \cap A^* S A^* = \emptyset.$$

In order to prove this lemma, we recall the notion of the *binomial coefficient* of two words $u, v \in A^*$. It is the number of occurrences of v as a subword of u and it is denoted by $\binom{u}{v}$. The binomial coefficient is completely characterized by the relations

$$\binom{ua}{vb} = \binom{u}{vb} + \binom{u}{v}\delta_{ab} \quad \text{for all } u, v \in A^*, a, b \in A,$$

$$\binom{u}{v} = 0 \quad \text{for all } u, v \in A^* \text{ with } |u| < |v|,$$

$$\binom{u}{1} = 1 \quad \text{for all } u \in A^*,$$

where $\delta_{ab}=1$ if $a=b$, $\delta_{ab}=0$ otherwise. Moreover one has

$$\binom{uw}{v}=\sum_{\substack{v_1,v_2\in A^*\\ v_1v_2=v}}\binom{u}{v_1}\binom{w}{v_2}\quad \text{for all } u, v, w\in A^*$$

(cf. [8]).

Let us now define the function $F: A^*\to\mathbb{N}$ by setting

$$F(w)=\sum_{ab\in S}\binom{w}{ab},\quad w\in A^*.$$

By means of the previous relations one can derive that for all $u_1, u_2\in A^*$, $a, b\in A$ one has

$$\begin{aligned} F(u_1abu_2)&=F(u_1bau_2)+1 \quad \text{if } ab\in S,\\ F(u_1abu_2)&=F(u_1bau_2) \qquad\ \ \text{if } ab\notin S. \end{aligned} \tag{6.1}$$

Proof of Lemma 6.6. Choose $v\in L_w(\pi)$ such that $F(v)$ is minimal. The word v is such that $L_v(\pi)\cap A^*SA^*=\emptyset$. Otherwise, indeed, one would have

$$w\overset{*}{\Rightarrow} v\overset{*}{\Rightarrow} u_1abu_2\Rightarrow u_1bau_2$$

for some $u_1, u_2\in A^*$, $ab\in S$ and therefore, in view of (6.1)

$$F(v)\geq F(u_1abu_2)>F(u_1bau_2)$$

which contradicts the minimality of $F(v)$. □

Proof of Proposition 6.5. Suppose that $L_2(A,\pi)$ is infinite. Then for all $n\geq 0$, there exists $w\in A^*$ such that $w\in L_2(A,\pi)$ and $|w|>n$. Choose v as in the previous lemma. Then one has

$$\begin{aligned} &|v|=|w|>n,\\ &L_v(\pi_2)\subseteq L_v(\pi)\subseteq L_w(\pi)\subseteq L_2(A),\\ &L_v(\pi_2)\cap A^*SA^*=\emptyset. \end{aligned}$$

We deduce that the projection of v into $M(A,\pi_2)$ is square-free and belongs to $M(A,\pi_2)\backslash J$.

Conversely, supppose that $M(A,\pi_2)$ contains infinitely many square-free elements. Then there are arbitrarily long words $w\in A^*$ such that

$$L_w(\pi_2)\subseteq L_2(A) \quad\text{and}\quad L_w(\pi_2)\cap A^*SA^*=\emptyset.$$

From the second of these equations we deduce $L_w(\pi)=L_w(\pi_2)$ and therefore one has $L_w(\pi)\subseteq L_2(A)$, that is $w\in L_2(A,\pi)$. By the arbitrariness of $|w|$ we conclude that $L_2(A,\pi)$ is infinite. □

In [4] it was shown that if π' is a symmetric system of semi-commutations on A such that $L_2(A,\pi)$ is infinite then there exists an infinite inalterable square-free word. A similar result holds if π is anti-symmetric (that is, if $\pi\cap\pi^{-1}$ is included in the identity relation).

Corollary 6.7. *Let π be an anti-symmetric system of semi-commutations on the alphabet A. Then $L_2(A, \pi)$ is infinite if and only if there exists an infinite inalterable square-free word.*

Proof. With the notations of Proposition 6.5, we have $\pi = \pi_1$, $S = \{ab \mid (ab, ba) \in \pi\}$ and for any $w \in A^*$, $L_w(\pi_2) = \{w\}$. Then $L_w(\pi_2)$ avoids S and the squares if and only if it is inalterable and square-free. The conclusion follows by Proposition 6.5. $\square$

We are not able to say whether there exists a system of semi-commutations π such that $L_2(A, \pi)$ is infinite and no infinite inalterable square-free word exists. Obviously such a system should be neither symmetric nor anti-symmetric.

Example 6.1. Let π be a system of semi-commutations on a three letter alphabet A. Then $L_2(A, \pi)$ is infinite if and only if π is included in the identity relation. Indeed if π contains a pair (ab, ba) with $a \neq b$, then $L_2(A, \{(ab, ba)\})$ is finite by Proposition 4.4 and, a fortiori, $L_2(A, \pi)$ is also finite.

Example 6.2. Consider the systems of semi-commutations π on a four letter alphabet A whose non-commutation graphs are shown in Fig. 2. In all these cases $L_2(A, \pi)$ is infinite. This is a consequence of Proposition 6.1 in cases (a) and (c) while in case (b) the infinite square-free word on four letters of Morse and Hedlund [9] is inalterable.

Cases (a) and (b) are "minimal" in the sense that $L_2(A, \sigma)$ is finite for all systems of semi-commutations σ properly containing π. For instance, in case (a), if one adds a new commutation to π then A becomes reducible. The restrictions of the non-commutation graph to the strongly connected components are shown in Fig. 3. We deduce by Proposition 6.3 and Example 6.1 that $L_2(A, \sigma)$ is reducible.

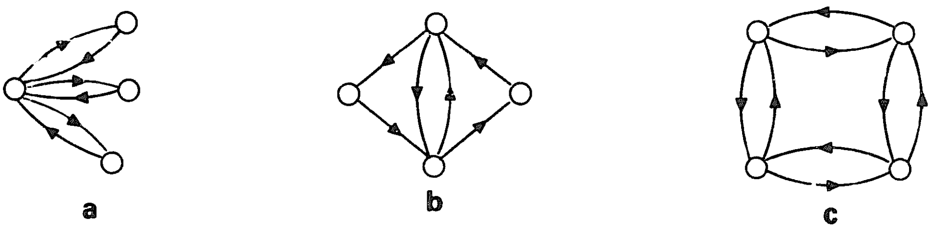

Fig. 2.

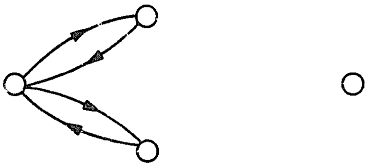

Fig. 3.

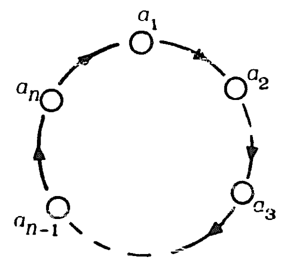

Fig. 4.

Example 6.3. Let π be a system of semi-commutations on the alphabet $A = \{a_1, a_2, \ldots, a_n\}$ whose non-commutation graph is shown in Fig. 4. Then one can prove by Proposition 6.5 that $L_2(A, \pi)$ is finite.

This example shows that there are arbitrarily large irreducible alphabets A such that $L_2(A, \pi)$ is finite. This seems to be an essential difference between the symmetric and the asymmetric case: indeed, if π is symmetric, A is irreducible and $\mathrm{Card}(A) \geq 5$, then $L_2(A, \pi)$ is infinite as shown in [4].

7. Overlap-free words and semi-commutations

A word $w \in A^*$ is said to be *overlap-free* if one cannot factorize w as $w = rasasat$ with $r, s, t \in A^*$ and $a \in A$. The set of the overlap-free words on the alphabet A will be denoted by $L_2'(A)$. Thue [10] has proved that $L_2'(A)$ is infinite whenever $\mathrm{Card}(A) \geq 2$.

Let π be a rewriting system on A. We say that a word $v \in A^*$ is *overlap-free relative to* π if $L_v(\pi) \subseteq L_2'(A)$. We denote by $L_2'(A, \pi)$ the set of overlap-free words relative to π.

Let us now suppose that π is a system of semi-commutations on A. In this case the "infinite overlap-free word problem" is recursively solvable. Indeed one has the following.

Proposition 7.1. *Let π be a system of semi-commutations on the alphabet A. Then $L_2'(A, \pi)$ is infinite if and only if the non-commutation graph contains a loop.*

Proof. Let us suppose that $L_2'(A, \pi)$ is infinite and show that the non-commutation graph contains a loop.

Since Corollary 6.4 can obviously be extended to overlap-free words, we can suppose, without loss of generality, that the non-commutation graph is strongly connected. Moreover, A must contain at least two letters and therefore the non-commutation graph contains a loop.

The converse is a straightforward consequence of the following lemma, since if $(a_1, a_2)(a_2, a_3) \ldots (a_n, a_1)$ is a loop in the non-commutation graph, then the word s defined there, is inalterable. $\square$

Lemma 7.2. *Let t be an infinite overlap-free word on the alphabet $\{a_1, a_2\}$ and s the word obtained from t by substituting the word $a_2a_3 \ldots a_na_1$ ($n \geq 2$) to each occurrence of the factor a_2a_1. Then s is an infinite overlap-free word on the alphabet $\{a_1, a_2, \ldots, a_n\}$.*

Proof. By contradiction. Suppose

$$s = xa_iya_iya_i \ldots \tag{7.1}$$

($1 \leq i \leq n$, $x, y \in A^*$). We can reduce ourselves to the case $i \leq 2$. Indeed, if $i > 2$, then each occurrence of a_i is preceded by $a_2a_3 \ldots a_{i-1}$, and therefore one has

$$s = x'a_2(a_3 \ldots a_iy')a_2(a_3 \ldots a_iy')a_2a_3 \ldots a_i \ldots$$

for suitable x', $y' \in A^*$. Let us then suppose $i \leq 2$. By deleting in (7.1) all the occurrences of $a_3, a_4, \ldots, a_n$ one obtains

$$t = \bar{x}a_i\bar{y}a_i\bar{y}a_i \ldots$$

($\bar{x}, \bar{y} \in \{a_1, a_2\}^*$). This is a contradiction, since t is overlap-free. □

References

[1] J. Berstel, Some recent results on square-free words, in: Lecture Notes in Computer Science **166** (Springer, Berlin, 1984) 14-25.

[2] A. Bertoni, M. Brambilla, G. Mauri and N. Sabadini, An application of the theory of free partially commutative monoids: Asymptotic densities of trace languages, in: Lecture Notes in Computer Science **118** (Springer, Berlin, 1981) 205-215.

[3] R.V. Book and C.C. Squier, Almost all one-rule Thue systems have decidable word problems, *Discrete Math.* **49** (1984) 237-240.

[4] A. Carpi and A. de Luca, Square-free words on partially commutative free-monoids, *Inform. Process. Lett.* **22** (1986) 125-131.

[5] P. Cartier and D. Foata, *Problèmes Combinatoires de Commutation et Réarrangements*, Lecture Notes in Mathematics **85** (Springer, Berlin, 1969).

[6] R. Cori and D. Perrin, Sur la reconnaissabilité dans les monoïdes partiellement commutatifs libres, *RAIRO Inform. Théor.* **19** (1985) 21-32.

[7] R. Dean, A sequence without repeats on x, x^{-1}, y, y^{-1}, *Amer. Math. Monthly* **72** (1965) 383-385.

[8] M. Lothaire, *Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **17** (Addison-Wesley, Reading, MA, 1983).

[9] M. Morse and G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.* **11** (1944) 1-7.

[10] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiania* **1** (1912) 1-67.